ChinaXiv合作期刊

干旱区地理
ARID LAND GEOGRAPHY

doi:10.12118/j.issn.1000-6060.2019.05.16

基于 *NDVI* 时间序列影像的天山北坡经济带农业种植结构提取①

熊元康[1,2]， 张清凌[1,3]

(1 中国科学院新疆生态与地理研究所,新疆 乌鲁木齐 830011； 2 中国科学院大学,北京 100049；
3 中山大学航空航天学院,广州 510006)

摘 要: 水资源匮乏是干旱区实现可持续发展的最大障碍。干旱区农业灌溉耗费大量的水资源,不同农作物在生长期所需的灌溉水量存在较大的差异,因此快速准确地了解干旱区的农业种植结构可以为节水型农业种植结构优化提供重要依据。以天山北坡经济带为研究区,以谷歌地球引擎(Google Earth Engine,GEE)云平台为支撑,以 Sentinel-2 以及 Landsat 7-8 的数据为遥感数据源,采取以下步骤进行研究区的农业种植结构提取:首先,为了简化农业种植结构提取的过程,利用一年最大 *NDVI* 值以及坡度信息构建耕地掩膜图层;然后,根据研究区内主要农作物的物候历,获取不同时间段内的最大 *NDVI* 值的时间序列数据以及农作物在一年中出现 *NDVI* 最大值的日期,并在此基础上构建一个包含 10 波段的特征波段影像;最后,结合野外实地考察获得的有效样本点以及经耕地掩膜图层掩膜后的 10 波段特征波段影像,利用随机森林分类器进行研究区的农业种植结构提取。分类结果表明:2018 年研究区内棉花、玉米、小麦的总体分类精度为 92.19%,Kappa 系数为 0.883。为了进一步将分类结果与统计数据进行对比,我们将训练得到的分类器应用于 2017 年的遥感影像,提取了研究区内 2017 年的农业种植结构信息,其分类结果表明 2017 年研究区内棉花、玉米、小麦的种植面积分别为 5 270 km^2、2 000 km^2、2 340 km^2,其相对精度分别为 86.53%、77.54%、86.19%。

关 键 词: 农业遥感； 谷歌地球引擎； 随机森林分类； 种植结构提取； 绿洲农业

农业种植结构是农作物空间格局的主要组成部分,指一个地区或生产单位内的农作物种植类型及其空间分布[1]。实时、准确地掌握农业种植结构的动态变化信息,可以为区域农作物估产、农业种植结构调整及其优化提供重要依据[2-4]。干旱区农业属于典型的灌溉农业,水资源匮乏是干旱区实现可持续发展的最大障碍,因此快速、准确地了解干旱区的农业种植结构可以为干旱区的节水型农业种植结构优化提供重要依据。由于利用人力调查一个地区或生产单位内的农业种植结构情况十分耗时耗力,近些年来遥感技术因其监测范围广、实时性强以及使用成本低等诸多优点,已被广泛地应用于对地观测中,其为快速和准确地获取大范围的农业种植结构信息提供了新的技术手段[5-6]。

目前,基于遥感的农业种植结构提取方法主要利用农作物的光谱特征、时相特征以及空间特征,概括起来主要有 3 大类:(1) 基于单一影像的农业种植结构提取方法。该方法适用于种植结构相对单一的地区,通过找到不同农作物的关键物候识别期以及农作物与其他地类在光谱特征上的差异进行农业种植结构提取。如 MATHUR 等[7]使用 IRS-1D 影像(Indian Remote Sensing satellite-1D)以及棉花和水稻的光谱特征,结合支持向量机(SVM)分类算法,成功提取了印度旁遮普区域内水稻和棉花的空间分

① **收稿日期:** 2019-01-14； **修订日期:** 2019-04-22
基金项目: 科技部国家重点研发计划(YFB0504204);中国科学院百人计划(Y674141001);新疆维吾尔自治区引进高层次人才专项(Y644151001)
作者简介: 熊元康(1994-),男,硕士研究生,主要从事农业遥感研究. E-mail xiongyuankang16@mails.ucas.ac.cn
通讯作者: 张清凌(1974-),男,研究员,研究方向为农业遥感、夜光遥感、城市遥感等. E-mail:zhangqling@mail.sysu.edu.cn

chinaXiv:201909.00154v1

ChinaXiv合作期刊

布状况。(2) 基于时间序列影像的农业种植结构提取方法。该方法充分利用农作物的季相特征,根据农作物在不同时期的生长状态来进行农作物识别。目前该方法已经成为农业种植结构提取的主流方法。如ZHONG等[8]采用不同植被指数(*EVI*,*NDSVI*等)的时序特征以及累积积温(Growing-degree-day,GDD)等特征,使用随机森林分类算法成功提取了美国堪萨斯州多尼芬县内玉米与大豆的空间分布状况。(3) 基于遥感影像与其他数据融合的农业种植结构提取方法。由于遥感数据受到获取能力等因素的影响,往往不能满足国家以及全球尺度下的农业种植结构提取。因此,许多学者尝试了将遥感数据与其他数据(如统计数据)进行融合的方式来实现国家以及全球尺度下的农业种植结构提取,如PORTMANN等[9]通过结合灌溉农作物和雨养农作物的物候历、农作物面积以及地形等信息进行重分类,成功获取了全球2000年5分栅格尺度的26种灌溉和雨养农作物的空间分布。

目前,利用时间序列影像进行大范围的农业种植结构提取主要使用的是低分辨影像,如MODIS数据,因其覆盖范围广、重复周期短等优势被广泛地应用于大尺度(区域、国家、全球)的农业种植结构提取,但影像的低分辨也会导致大量混合像元产生,从而降低农业种植结构提取的精度[10-11]。中高分辨率影像能减少混合像元,但由于其监测范围小和重返周期长等因素限制了其在大范围内进行农业种植结构提取的应用。目前,使用中高分辨率遥感影像在干旱区进行大范围的农业种植结构提取的研究较少。为解决大范围的干旱区农业种植结构提取所面临的低分辨率遥感影像混合像元严重和中高分辨率遥感影像监测范围小以及重返周期长等问题,本文以天山北坡经济带为研究区,基于Google Earth Engine(GEE)云平台,结合使用中高分辨率的Sentinel-2和Landsat 7-8等多源遥感数据,通过构建*NDVI*时间序列以及其他辅助特征进行天山北坡经济带的农业种植结构快速提取。

GEE云平台是一个能在大尺度范围下进行遥感数据处理以及分析的云平台,该平台提供了一个完备的集成环境,其中包括遥感数据库(如Sentinel-2A/B、Landsat系列、MODIS等遥感数据),以及建立在JavaScript(或Python) APIs上的交互式算法开发环境[12]。目前,国内外多个研究组基于该平台开展了各种对地观测研究,如水稻遥感制图[13],耕地遥感提取[14],城市变化遥感检测[15],冬小麦遥感制图[16]等。

1 研究区概况与数据源

1.1 研究区概况

天山北坡经济带地处亚欧大陆腹地,位于天山北麓、准噶尔盆地南缘,属于干旱与半干旱气候区,地理范围为79°52′~91°32′E,43°01′~46°13′N。该区域以乌鲁木齐市为中心,东起木垒哈萨克自治县,西至温泉县,横跨17个县市(图1)[17]。近十几年来,该地区每年的国内生产总值占新疆维吾尔自治区国内生产总值的50%左右,是新疆经济、教育、科技等最为发达的地区,也是丝绸之路经济带通往中亚和欧洲的重要枢纽[17-18]。但由于受干旱气候环境的影响,水资源匮乏已成为制约该地区实现可持续发展的最大因素[19],而该地区的农业耗水占该地区总耗水的91%[20]。因此,深入了解该地区的农业种植结构,有利于该地区的水资源优化配置以及可持续发展。根据历年的《新疆统计年鉴》可知,该地区主要种植的农作物为棉花、玉米以及小麦等一年一熟的农作物,2017年这3类农作物的种植面积占整个地区农作物种植面积的76.9%[21]。

1.2 数据源

1.2.1 遥感数据 多时相的遥感数据能够充分地反映农作物的生长特征[22-24],本文根据不同农作物在不同时期的生长状况信息(表1)来进行研究区的农业种植结构提取。为了充分获取研究区内不同农作物在不同时期的生长状况信息,在农作物种植结构提取的识别期内,我们在GEE云平台上一共获得了4 313景遥感影像,其中Sentinel-2影像3 582景,Landsat-8影像381景,Landsat-7影像350景(表2)。综合以上多源遥感影像一方面可以增加可用的遥感数据(光学遥感影像往往容易受到云的干扰使得可用像元减少)。另一方面,由于各卫星过境时间不同,融合使用多源遥感数据可以提高遥感监测时间频率以更好地捕捉农作物的生长过程(如Landsat-8在2018年3月在新疆过境的时间分别为3月7日与3月23日,而Landsat-7在2018年3月份在新疆过境的时间分别为3月15日与3月31日)。

1.2.2 地面调查数据 我们于2018年5~6月在研究区内进行了野外实地考察,一共记录了有效样本点822个。其中,棉花有效样本点394个,玉米

chinaXiv:201909.00154v1

ChinaXiv合作期刊

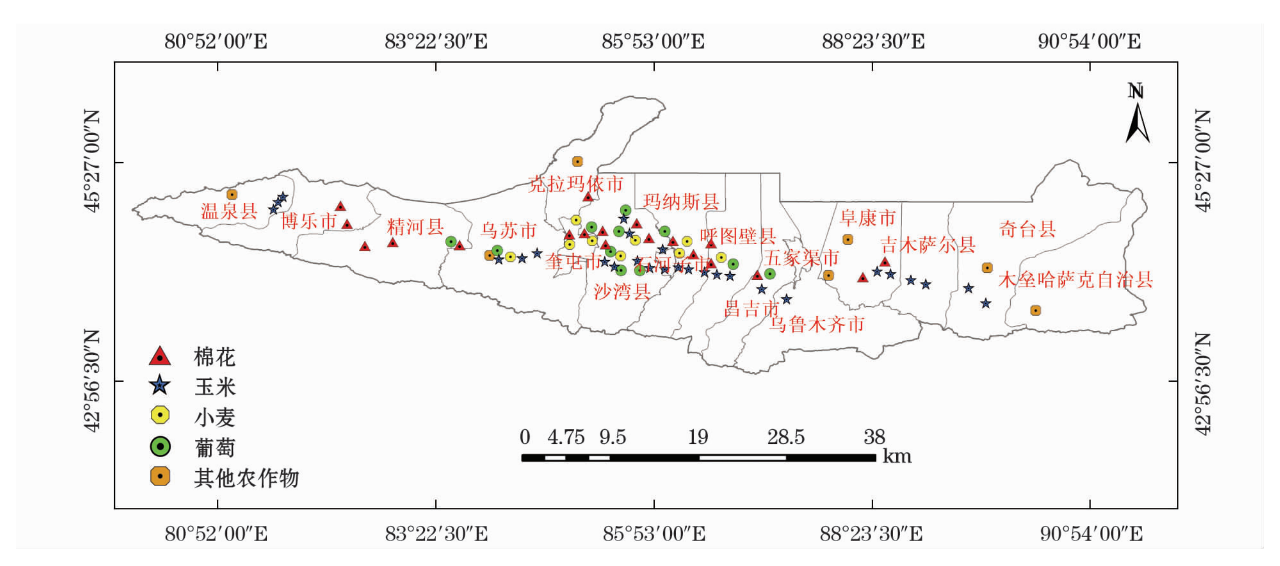


图 1　研究区位置与地面调查采样点

Fig. 1　Location of the study area and the spatial distribution of the samples

表 1　研究区内主要农作物的物候历

Tab. 1　Crops calendars in study area

月	6 月			前一年 9 月			7 月			3 月			8 月			4 月			9 月			5 月		
旬	上	中	下	上	中	下	上	中	下	上	中	下	上	中	下	上	中	下	上	中	下	上	中	下
棉花							播种			出苗		苗期			蕾期			花铃期				吐絮期		
冬小麦	播种							拔节			抽穗			乳熟		成熟								
春小麦				播种				拔节			抽穗		乳熟		成熟									
春玉米								播种			出苗		苗期			拔节		抽雄				乳熟		

表 2　实验选取的遥感影像源

Tab. 2　Remote sensing data sources used in this study

传感器类型	获取时间	波段	波段名称	波长/nm	分辨率/m	数据源
Sentinel - 2 MSI,TOA	2018 - 03 - 01—2018 - 12 - 01	B4	Red	665	10	ESA
		B8A	NIR	842	10	
Landsat 8 OLI,TOA	2018 - 03 - 01—2018 - 12 - 01	B4	Red	640 ~ 670	30	USGS
		B5	NIR	850 ~ 880	30	
Landsat7 ETM + ,TOA	2018 - 03 - 01—2018 - 12 - 01	B3	Red	630 ~ 690	30	USGS
		B4	NIR	770 ~ 900	30	
Shuttle Radar Topography Mission (*SRTM*)		Slope			30	USGS

有效样本点 220 个,小麦有效样本点 67 个,葡萄有效样本点 51 个,其他农作物(包括打瓜、辣椒、葫芦瓜等)有效样本点 90 个。在进行地面采样的过程中,我们保证了各个样本点之间的独立性,即样本点之间均为独立的单个像元,以增强分类的可靠性,有效样本点在研究区内的空间分布如图 1 所示。

1.2.3　统计数据　根据最新的《新疆统计年鉴》(2017 年),我们统计了该区域内各主要农作物(棉花、玉米、小麦等)的种植面积,并将其作为评价标准,对遥感提取的各主要农作物的种植面积进行精度评价。

1.3　数据预处理

云以及云阴影对光学遥感影像具有显著的影响[25],它们会严重影响基于遥感影像所进行的研究,如:遥感影像合成[26]、植被指数计算[27]以及地表覆盖分类[28]等。我们在 GEE 云平台上采用了

chinaXiv:201909.00154v1

ChinaXiv合作期刊

chinaXiv:201909.00154v1

FMASK[29]算法对获取的遥感数据进行了去云以及去云阴影处理,以减少数据合成,植被指数计算等对农业种植结构提取所造成的噪声影响。同时将Sentinel-2的数据产品(10 m)重采样至Landsat 7-8的30 m分辨率,以保证使用的遥感数据在分辨率上的一致。

2 研究方法

2.1 耕地掩膜图层构建

为了减少非农作物植被对农业种植结构提取的影响,我们通过以下步骤提取了研究区2018年的耕地掩膜图层:首先,我们根据Google Earth的高清影像以及实地调研情况,在研究区内选择了大量的非农作物植被样本点,如森林、天然草地等,将收集到的非农作物植被样本点与有效的农作物样本点在各类遥感数据(如:*NDVI*、DEM等数据产品)上进行分析对比;其次,根据分析对比的结果以及以往的研究,构建耕地掩膜判别规则:$NDVI_{Max} > 0.4$[30]和$Slope < 7°$。其中$NDVI_{Max}$代表一年中最大的*NDVI*值,*Slope*代表坡度信息,由SRTM数据计算所得;最后,以Landsat-8的遥感数据以及SRTM数据为数据源,根据构建的规则提取耕地掩膜图层,并将获得的耕地掩膜图层与Google Earth的高清影像进行叠加,通过人工目视解译去除一些明显提取错误的区域,最终获得研究区2018年30 m分辨率的耕地掩膜图层(图2)。为了检验耕地提取的精度,我们在生成的耕地掩膜图层中随机生成1 000个样本点,并利用Google Earth的高清影像对生成的样本点进行人工目视解译以判断其是否属于耕地。结果表明在这1 000个随机样本点中有967个为耕地样本点,33个为非耕地样本点(其中多为防护林以及其他自然植被),耕地提取的精度为96.7%,可以满足下一步的要求。

2.2 农作物分类特征构建

2.2.1 *NDVI*时序特征构建 植被指数常常被作为特征参数来评估地表植被的生长状况,其中归一化植被指数(Normalized Difference Vegetation Index, *NDVI*)使用最为广泛[31-33]。*NDVI*表达了红波段(植物吸收强烈)与近红外波段(植被反射强烈)之间的关系,其计算公式如下所示[34]:

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}} \tag{1}$$

式中:ρ_{NIR}代表近红外波段的反射率;ρ_{Red}代表红波段的反射率。由于*NDVI*能够很好地反映植被的生长状况,因此本文将*NDVI*作为评估农作物在不同时期生长状况的主要特征参数。

不同传感器在设计上的差异,导致了不同传感器具有不同的光谱响应函数(Spectral Response Functions, *SRF*)。比如,Sentinel-2与Landsat-7的数据在红波段的均方根误差(Root Mean Square Error, *RMSE*)超过了8%[35]。为了减小由于不同传感器的不同光谱响应函数所带来的误差,我们以Landsat-8数据计算的*NDVI*值为基准,将Sentinel-2以及Landsat-7数据计算所得的*NDVI*值进行相关的线性

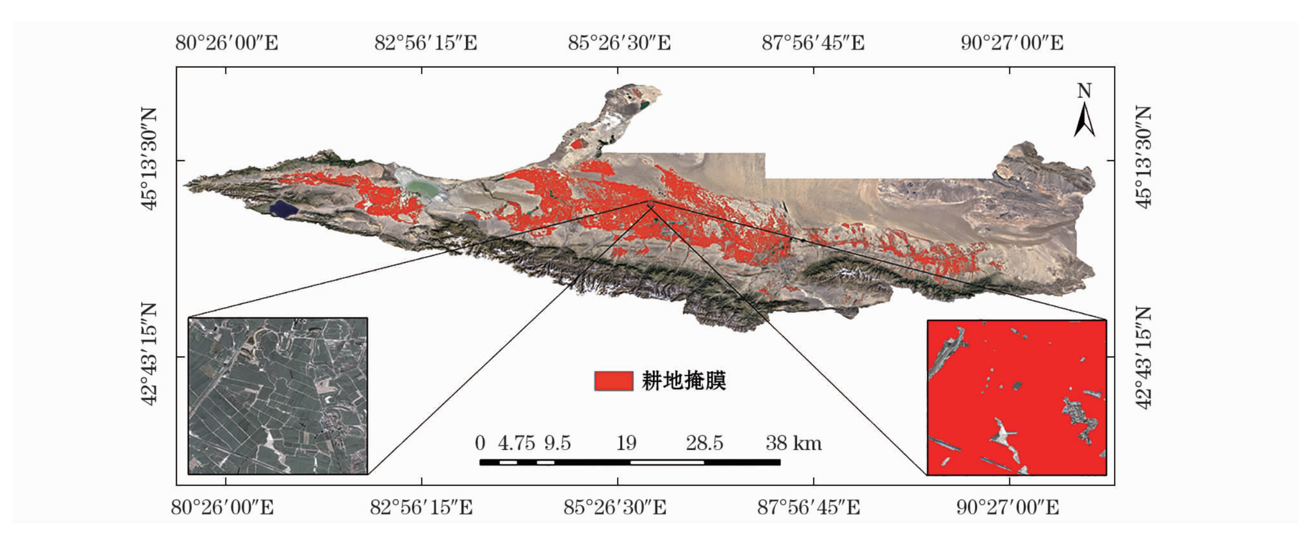


图2 研究区的耕地提取结果

Fig. 2 Extracted cropland in the study area

ChinaXiv合作期刊

转换。其采用的线性模型计算公式如下所示[36－37]：

$$NDVI_{\text{Landsat-8}} = 0.0235 + 0.9723 \times NDVI_{\text{Landsat-7}} \quad (2)$$

$$NDVI_{\text{Landsat-8}} = -0.0046 + 1.0183 \times NDVI_{\text{Sentinel-2}} \quad (3)$$

根据研究区内主要农作物的物候识别期（表1），我们利用去云后的 Sentinel-2 以及 Landsat 7-8 数据计算了2018 年3 月1 日～2018 年12 月1 日这个时间段内的所有 *NDVI* 值，同时根据以上的线性模型进行了 *NDVI* 值的转换，并依据获取影像的时间对 *NDVI* 数据进行时间排序。与此同时，我们对排好序的 *NDVI* 时间序列，计算 2018 年 3 月 1 日～2018 年 12 月 1 日这个时间段内每个月的最大 *NDVI* 值，并将每个月的最大 *NDVI* 值作为一个特征波段。

2.2.2　辅助特征构建　通过对研究区内不同农作物的物候历进行深入分析后，我们发现不同农作物出现生长峰值的时间是有差异的。比如棉花出现 *NDVI* 最大值的日期在每年的第 225 d 左右，玉米出现 *NDVI* 最大值的日期在每年的第 195 d 左右，小麦出现 *NDVI* 最大值的日期在每年的第 155 d 左右（图3a）。为了更加突出不同农作物在不同时期的生长状况，我们将农作物在一年中获得 *NDVI* 最大值的日期也作为分类特征参数。为了扩大最大 *NDVI* 值与出现最大 *NDVI* 值日期的差异，我们采用以下公式将其进行归一化处理：

$$Normalize_Date = \frac{Sum_{day} - Max_{day}}{Sum_{day}} \quad (4)$$

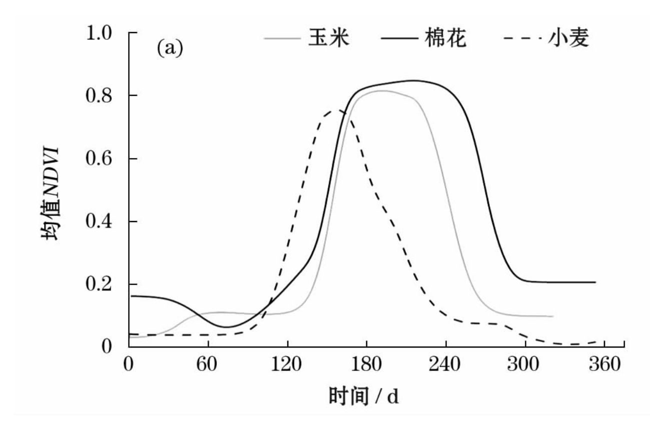
式中：Sum_{day}代表该年一共有多少天，Max_{day}代表一年中出现 *NDVI* 最大值的日期。如小麦 *NDVI* 出现最大值的日期约为第 155 d（5 月份左右），其 *Normalize_Date* 值则为 0.57，而棉花 *NDVI* 出现最大值的日期约为第 255 d（7 月份左右），其 *Normalize_Date* 值则为 0.3，既 *NDVI* 最大值出现的越早，其 *Normalize_Date* 值越大。

根据 2.2.1 获得的 *NDVI* 时相特征以及 *NDVI* 最大值日期特征，我们一共获取了 10 个分类特征量（F1：Max_$NDVI_{Mar}$；F2：Max_$NDVI_{Apr}$；F3：Max_$NDVI_{May}$；F4：Max_$NDVI_{Jun}$；F5：Max_$NDVI_{Jul}$；F6：Max_$NDVI_{Aug}$；F7：Max_$NDVI_{Seq}$；F8：Max_$NDVI_{Oct}$；F9：Max_$NDVI_{Nov}$；F10：Normalize_Date）。研究区内主要农作物的分类特征曲线如图 3b 所示。

2.3　随机森林分类器

随机森林分类（Random Forest，RF）是一种集成学习分类器，是由一系列 CART 决策树以及投票机制组合而成的机器学习算法[38]。它比其他分类器在进行遥感图像分类时表现得更加健壮，分类速度相对更快，更容易实现[36]。它使用 bootstrap 抽样技术从原始数据集中抽取训练集，并通过训练集构建 CART 决策树。在构建 CART 决策树的过程中，同时采取一些规则（如 Gini 系数最小原则）来计算最佳分割方式并进行内部节点分割，以便减少 CART 决策树之间的相关性，减少过拟合现象。

我们将在 GEE 云平台上集成的随机森林算法应用于本文的研究[12]。GEE 云平台上的随机森林分类器的参数如下所述：（1）随机森林分类器的决策树数量；（2）候选特征子集；（3）叶节点最小样本数；（4）再抽样比率；（5）是否计算袋外错误率（out-of-bag，OOB）；（6）随机种子数。其中随机森林分类器的决策树数量、候选特征子集以及叶节点最小样本数是影响随机森林分类器性能与效率的最主要参数。我们根据已有的大量研究[39-41]以及本



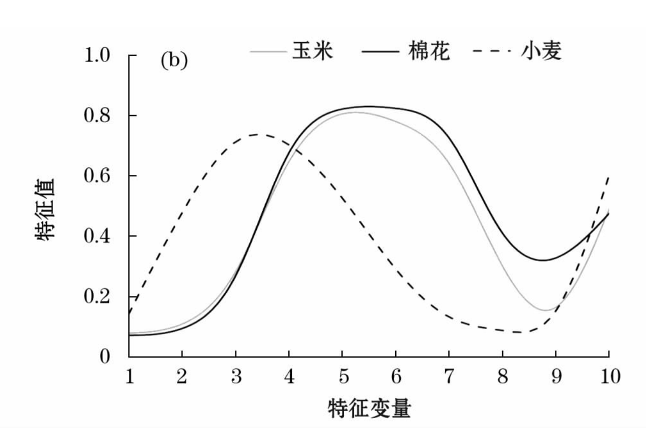


图 3　研究区内主要农作物的 *NDVI* 值以及分类特征值变化曲线

Fig. 3　Time-series of *NDVI* and classification variables of major crops in the study area

chinaXiv:201909.00154v1

ChinaXiv合作期刊

文研究的实际情况,将以下 3 个参数分别设置为:随机森林分类器的决策树数量 = 100;候选特征子集 = 4;叶节点最小样本数 = 1。根据在 2.2 中构建的分类特征,在 GEE 云平台上融合生成了 10 波段的分类特征影像,并将野外实地考察获得的有效样本点随机选择一半作为训练集进行随机森林分类器的训练,另一半作为测试集验证其分类精度。

3 分析与讨论

3.1 精度分析

本文采用混淆矩阵精度以及与统计数据对比这两种方式对分类结果进行精度验证。由于随机森林分类器是基于像元的分类器,分类结果中往往存在椒盐噪声。在进行混淆矩阵计算之前,我们采用众数滤波(3 ×3 pixels)对分类结果进行滤波处理。最终用计算得到的用户精度(user's accuracy,*UA*)、生产者精度(produce's accuracy,*PA*)、总体精度(overall accuracy,*OA*)以及 Kappa 系数来表征分类结果的混淆矩阵精度(表 3)。如表 3 所示,分类结果的总体精度为 92.19%,Kappa 系数为 0.883,主要农作物的用户精度以及生产者精度都超过了 85%,但其他农作物的用户精度只有 64.44%。表明其他农作物(辣椒、打瓜、葵花等)也具有与棉花、玉米等作物相似的分类特征值,特别是易于与玉米混淆。

由于我们提取的是研究区内 2018 年主要农作物的种植结构信息,而目前对于研究区内的农业种植结构信息只统计到了 2017 年,因此我们将 2018 年农业种植结构提取所训练的模型应用于 2017 年研究区内的农业种植结构提取,以便与 2017 年的统计数据进行比较。据统计,2017 年研究区内棉花、玉米、小麦的种植面积分别为 4 650 km^2、1 630 km^2、2 050 km^2。

根据 2017 年的分类结果得知,2017 年研究区内棉花、玉米、小麦的种植面积分别为 5 270 km^2、2 000 km^2、2 340 km^2,其分类精度分别为 86.53%、77.54%、86.19%。同时我们还统计了各县市的农作物分类结果,并与统计数据进行了对比(图 4)。由于五家渠市与石河子市属于新疆生产建设兵团地区,统计年鉴中缺少这两个地区的农业种植结构统计信息,因此本文没有对这两个地区的分类结果进行精度分析。由图 4 可知,乌鲁木齐市、克拉玛依市、昌吉回族自治州这 3 个地区的分类结果较统计数据而言有较大的误差,其主要原因是这 3 个地区的种植结构相对于其他地区更为复杂,导致了其他农作物与这 3 类主要农作物产生了混淆。但由于 2017 年农业种植结构提取模型是基于 2018 年的野外实地考察数据训练获得的,而不同年份之间的农业种植制度是有所差异的,因此更加准确的精度验证结果仍需新的统计数据(2018 年)加以论证。

3.2 种植结构提取结果

本文利用随机森林分类器以及构建的分类特征,成功获取了研究区内 2018 年 30 m 分辨率的农业种植结构信息(图 5)。研究区内棉花种植主要分布在精河县、博乐县、乌苏市、克拉玛依市、奎屯市、沙湾县、石河子市、玛纳斯县、呼图壁县、五家渠市等 10 个地区;玉米种植主要分布在温泉县、博乐市、呼图壁县、昌吉市、阜康市、吉木萨尔县、奇台县、木垒县等 8 个地区;小麦种植主要分布在奇台县、昌吉市、沙湾县、奎屯市、乌苏市、温泉县等 6 个地区;葡萄种植主要分布在乌鲁木齐市、阜康市、呼图壁县、玛纳斯县等 4 个地区。同时我们将遥感反演的主要农作物的空间分布信息与历年《新疆统计年鉴》中各地区的种植结构信息以及实地考察信息进行对比后发现,遥感反演的主要农作物的空间分布能较好

表 3 分类精度

Tab. 3 Accuracy of the classification result

分类后数据	参考数据					总计	用户精度
	棉花	玉米	小麦	葡萄	其他农作物		
棉花	190	5	0	2	0	197	96.44%
玉米	2	106	0	0	2	110	96.39%
小麦	1	2	29	0	1	33	87.87%
葡萄	1	0	0	24	0	25	96.00%
其他农作物	4	10	1	1	29	45	64.44%
总计	198	123	30	27	32	总体精度 = 92.19%	
生产者精度	95.95%	86.17%	96.66%	88.89%	90.62%	Kappa 系数 = 0.883	

chinaXiv:201909.00154v1

ChinaXiv合作期刊

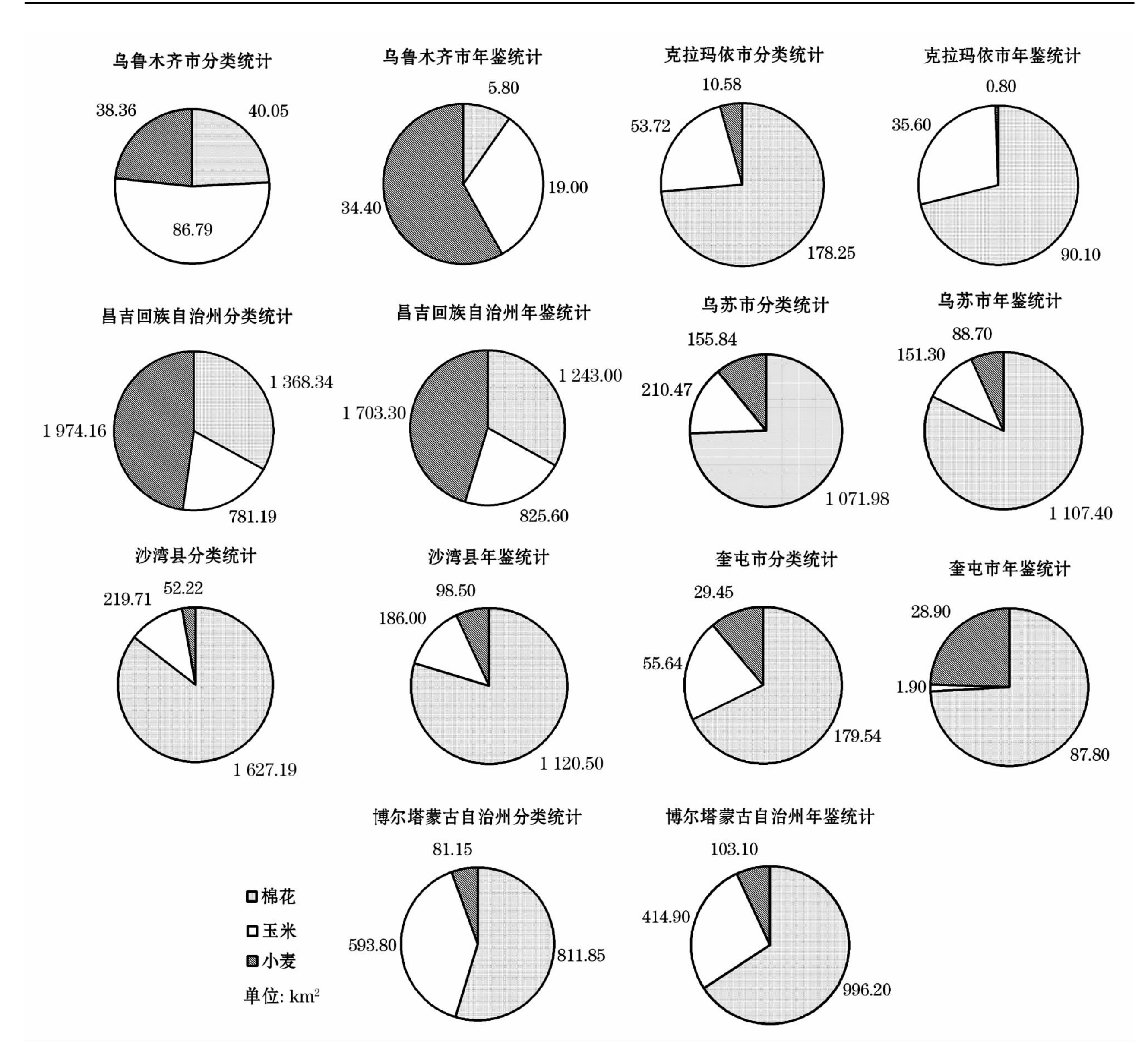


图 4　2017 年农作物分类结果与统计年鉴数据的比较

Fig. 4　Comparison between classification results and statistical reports in 2017

的与之相对应。

虽然根据遥感分类的结果来看,运用随机森林分类器以及构建的分类特征能较好的应用于研究区的农业种植结构提取,但是仍有以下问题需要进一步分析:由于天山北坡经济带主要从事的是 1 年 1 季的农业生产且种植结构较为单一,因此,本文并未考虑多季、套种、间种等农业种植模式;从遥感分类结果来看,棉花、玉米、小麦是研究区的主要农作物,这一点与统计年鉴的统计信息是基本吻合的。但由于其他作物(辣椒、葵花等)也具有与这 3 类主要农作物相似的季相节律,导致了遥感提取的这 3 类农作物的种植面积大于实际的种植面积,因此在后续的研究中需要更加深入地了解该地区各个农作物在其他特征(空间特征等)上表现出来的差异,以便提高分类的精度。

4　结　论

受制于有限的水资源,农业种植结构优化以及采用节水型农业灌溉技术是现代干旱区绿洲农业实现可持续发展的必经之路。本文以天山北坡经济带为研究区,通过利用研究区内主要农作物的物候历信息以及野外考察数据,基于遥感数据构建了 10 个分类特征。同时基于 GEE 云平台融合生成了研究区内 Sentinel-2、Landsat 7-8 的 *NDVI* 时序数据以及 1 年中出现*NDVI*最大值的日期数据,并在此基础上

chinaXiv:201909.00154v1

ChinaXiv合作期刊

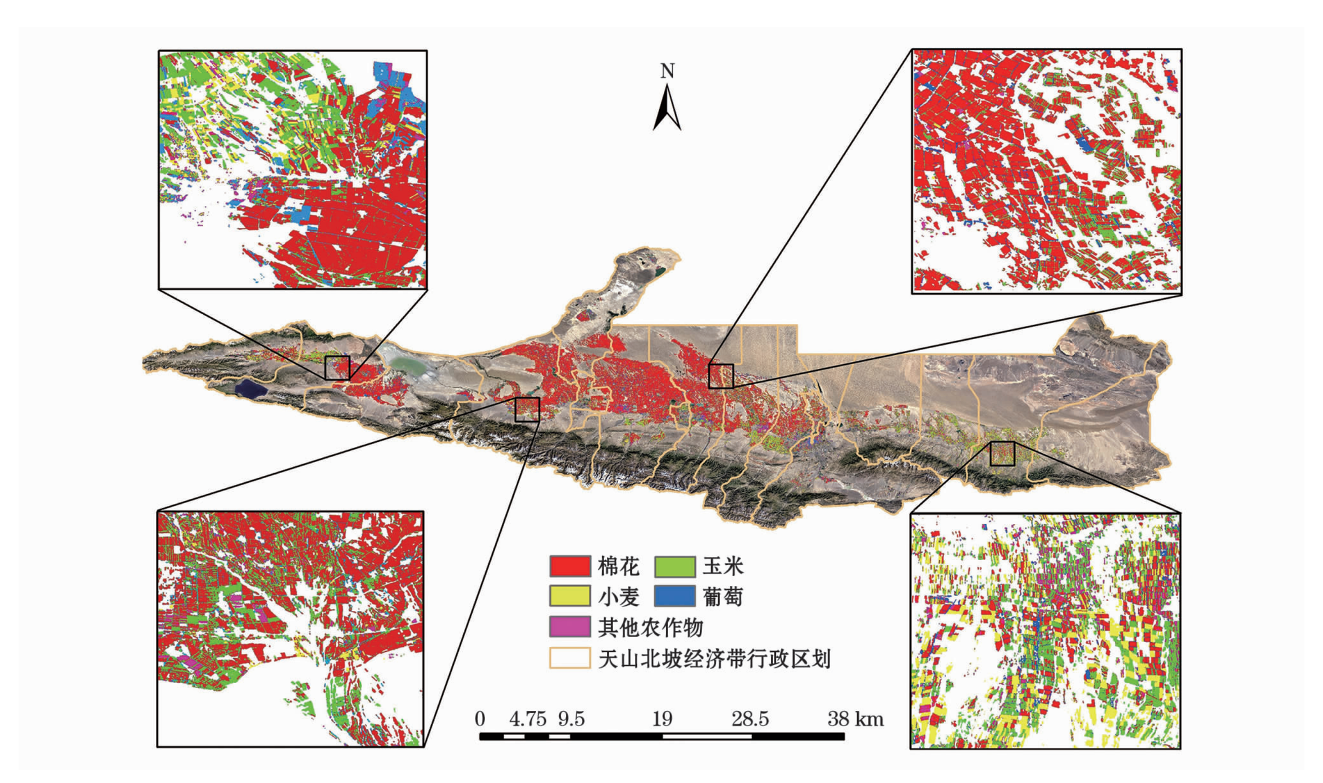


图 5　研究区内 2018 年主要农作物的空间分布

Fig. 5　Spatial distribution of major crops in the study area in 2018

构建了 10 波段的特征波段影像。同时通过野外考察获得的有效样本点以及随机森林分类器进行了研究区内 2018 年主要农作物(棉花、玉米、小麦)的种植结构提取。根据对其分类结果进行分析后,我们得出以下主要结论:

(1) 多源数据的融合使用能更加准确的表征不同农作物在不同时期的生长状态。由于我们综合使用了 Sentinel-2 以及 Landsat 7-8 作为遥感数据源,因此能更为准确地获取不同农作物在不同时期的生长状况以及 1 年中出现 *NDVI* 最大值的日期,同时构建的最大 *NDVI* 值特征又能够有效的去除薄云及其阴影对种植结构提取的影响。

(2) 构建的 10 个分类特征以及随机森林分类器能够较好的适应于研究区内的种植结构提取。运用构建的分类特征以及随机森林分类器进行种植结构提取的总体精度达到了 92.19%,Kappa 系数为 0.883。与统计数据对比也说明了构建的 10 个分类特征以及随机森林分类器能够较好的适应于该研究区的种植结构提取。

(3) 天山北坡经济带内主要种植的农作物为棉花、玉米、小麦,这 3 种农作物的播种面积占整个研究区农作物播种面积的 75% 左右。同时,新的对地观测卫星如 Sentinel-2 星座以及遥感云平台如 GEE 的出现,为进行基于时间序列影像的大区域农业种植结构快速提取带来了新的技术手段。

参考文献(References)

[1] 唐华俊,吴文斌,杨鹏,等. 农作物空间格局遥感监测研究进展[J]. 中国农业科学,2010,43(14):2879－2888. [TANG Huajun,WU Wenbin,YANG Peng,et al. Recent progresses in monitoring crop spatial patterns by using remote sensing technologies[J]. Scientia Agricultura Sinica,2010,43(14):2879－2888.]

[2] OZDOGAN M. The spatial distribution of crop types from MODIS data: Temporal unmixing using independent component analysis [J]. Remote Sensing of Environment, 2010, 114 (6): 1190－1204.

[3] 马丽,顾晓鹤,徐新刚,等. 地块数据支持下的玉米种植面积遥感测量方法[J]. 农业工程学报,2009,25(8):147－151. [MA Li,GU Xiaohe,XU Xingang,et al. Remote sensing measurement of corn planting area based on field-data[J]. Transactions of the CSAE,2009,25(8):147－151.]

[4] WARDLOW B D,EGBERT S L,KASTENS J H. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains[J]. Remote Sensing of Environment,2007,108(3):290－310.

[5] XIAO X,BOLES S,FROLKING S,et al. Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images[J]. Remote Sensing of Environment,2006,100(1):95－113.

[6] VINTROU E,DESBROSSE A,BÉGUÉ A,et al. Crop area mapping in West Africa using landscape stratification of MODIS time series

chinaXiv:201909.00154v1

ChinaXiv合作期刊

and comparison with existing global land products[J]. International Journal of Applied Earth Observation and Geoinformation,2012,14(1):83 – 93.

[7] MATHUR A,FOODY G M. Crop classification by support vector machine with intelligently selected training data for an operational application[J]. International Journal of Remote Sensing,2008,29(8):2227 – 2240.

[8] ZHONG L,GONG P,BIGING G S. Efficient corn and soybean mapping with temporal extendability:A multi-year experiment using Landsat imagery[J]. Remote Sensing of Environment,2014,140:1 – 13.

[9] PORTMANN F T, SIEBERT S, DÖLL P. MIRCA2000-Global monthly irrigated and rainfed crop areas around the year 2000:A new high-resolution data set for agricultural and hydrological modeling[J]. Global Biogeochemical Cycles,2010,24(1):1 – 24.

[10] JIA K,LIANG S,WEI X,et al. Land cover classification of Landsat data with phenological features extracted from time series MODIS NDVI data[J]. Remote Sensing,2014,6(11):11518 – 11532.

[11] GUSSO A,FORMAGGIO A R,RIZZI R,et al. Soybean crop area estimation by MODIS/EVI data[J]. Pesquisa Agropecuária Brasileira,2012,47(3):425 – 435.

[12] GORELICK N,HANCHER M,DIXON M,et al. Google earth engine:Planetary-scale geospatial analysis for everyone[J]. Remote Sensing of Environment,2017,202:18 – 27.

[13] DONG J,XIAO X,MENARGUEZ M A,et al. Mapping paddy rice planting area in Northeastern Asia with Landsat 8 images,phenology-based algorithm and Google Earth Engine[J]. Remote Sensing of Environment,2016,185:142 – 154.

[14] XIONG J,THENKABAIL P S,TILTON J C,et al. Nominal 30-m cropland extent map of continental Africa by integrating pixel-based and object-based algorithms using sentinel-2 and Landsat-8 data on Google Earth Engine[J]. Remote Sensing,2017,9(10):1065.

[15] GOLDBLATT R,YOU W,HANSON G,et al. Detecting the boundaries of urban areas in India:A dataset for pixel-based image classification in Google Earth Engine[J]. Remote Sensing,2016,8(8):634.

[16] 王九中,田海峰,邬明权,等. 河南省冬小麦快速遥感制图[J]. 地球信息科学学报,2017,19(6):846 – 853. [WANG Jiuzhong, TIAN Haifeng, WU Mingquan, et al. Rapid mapping of winter wheat in Henan Province[J]. Journal of Geo-information Science,2017,19(6):846 – 853.]

[17] 朱慧,王哲,焦广辉,等. 天山北坡经济带经济格局时空演变特征研究[J]. 干旱区资源与环境,2012,26(3):23 – 29. [ZHU Hui,WANG Zhe,JIAO Guanghui,et al. Spatial dynamic evolution of economic pattern in economic belt on the north slope of Tianshan[J]. Journal of Arid Land Resources and Environment,2012,26(3):23 – 29.]

[18] 董雯,杨宇,周艳时. 干旱区绿洲城市土地利用效益研究——以乌鲁木齐为例[J]. 干旱区地理,2011,34(4):679 – 685. [DONG Wen,YANG Yu,ZHOU Yanshi. Land use changes and evaluation of land use efficiency in an arid oasis city:A case of Urumqi,China[J]. Arid Land Geography,2011,34(4):679 – 685.]

[19] DINDA S, COONDOO D, PAL M. Air quality and economic growth:An empirical study[J]. Ecological Economics,2000,34(3):409 – 423.

[20] 吴业鹏,袁汝华. 丝绸之路经济带背景下新疆水资源与经济社会协调性评价[J]. 水资源保护,2016,32(4):60 – 66. [WU Yepeng, YUAN Ruhua. Evaluation of coordination between water resources and economic and social development in Xinjiang in context of Silk Road Economic Belt[J]. Water Resources Protection,2016,32(4):60 – 66.]

[21] 新疆维吾尔自治区统计局. 新疆统计年鉴[M]. 北京:中国统计出版社,2017. [Statistic Bureau of Xinjiang Uygur Autonomous Region. Xinjiang statistical yearbook[M]. Beijing:China Statistics Press,2017.]

[22] PAN Y,HU T,ZHU X,et al. Mapping cropland distributions using a hard and soft classification model[J]. IEEE Transactions on Geoscience and Remote Sensing,2012,50(11):4301 – 4312.

[23] CLAVERIE M,DEMAREZ V,DUCHEMIN B,et al. Maize and sunflower biomass estimation in Southwest France using high spatial and temporal resolution remote sensing data[J]. Remote Sensing of Environment,2012,124:844 – 857.

[24] LHERMITTE S,VERBESSELT J,VERSTRAETEN W W,et al. A comparison of time series similarity measures for classification and change detection of ecosystem dynamics[J]. Remote Sensing of Environment,2011,115(12):3129 – 3152.

[25] ZHU Z,WOODCOCK C E. Object-based cloud and cloud shadow detection in Landsat imagery[J]. Remote Sensing of Environment,2012,118:83 – 94.

[26] ROY D P, JU J, KLINE K, et al. Web-enabled Landsat Data (WELD):Landsat ETM + composited mosaics of the conterminous United States[J]. Remote Sensing of Environment,2010,114(1):35 – 49.

[27] HUETE A,DIDAN K,MIURA T,et al. Overview of the radiometric and biophysical performance of the MODIS vegetation indices[J]. Remote Sensing of Environment,2002,83(1 – 2):195 – 213.

[28] ZHANG Y,GUINDON B,CIHLAR J. An image transform to characterize and compensate for spatial variations in thin cloud contamination of Landsat images[J]. Remote Sensing of Environment,2002,82(2 – 3):173 – 87.

[29] ZHU Z,WANG S,WOODCOCK C E. Improvement and expansion of the Fmask algorithm:Cloud,cloud shadow,and snow detection for Landsats 4 – 7,8,and Sentinel 2 images[J]. Remote Sensing of Environment,2015,159:269 – 277.

[30] ZHANG Q,LI B,THAU D,et al. Building a better urban picture:Combining day and night remote sensing imagery[J]. Remote Sensing,2015,7(9):11887 – 11913.

[31] TOWNSHEND J R,GOFF T E,TUCKER C J. Multitemporal dimensionality of images of normalized difference vegetation index at continental scales[J]. IEEE Transactions on Geoscience and Remote Sensing,1985,6(6):888 – 895.

[32] TARPLEY J,SCHNEIDER S,MONEY R L. Global vegetation indices from the NOAA-7 meteorological satellite[J]. Journal of Climate and Applied Meteorology,1984,23(3):491 – 494.

[33] ROSENTHAL W D,BLANCHARD B J,BLANCHARD A J. Visible/infrared/microwave agriculture classification, biomass, and plant height algorithms[J]. IEEE Transactions on Geoscience and Remote Sensing,1985,23(2):84 – 90.

[34] TUCKER C J. Red and photographic infrared linear combinations

chinaXiv:201909.00154v1

ChinaXiv合作期刊

for monitoring vegetation[J]. Remote Rensing of Environment, 1979,8(2):127－150.

[35] D'ODORICO P,GONSAMO A,DAMM A,et al. Experimental evaluation of Sentinel-2 spectral response functions for NDVI time-series continuity[J]. IEEE Transactions on Geoscience and Remote Sensing,2013,51(3):1336－1348.

[36] ROY D, KOVALSKYY V, ZHANG H, et al. Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity[J]. Remote Sensing of Environment,2016,185:57－70.

[37] MANDANICI E,BITELLI G. Preliminary comparison of Sentinel-2 and Landsat 8 Imagery for a combined use[J]. Remote Sensing, 2016,8(12):1014.

[38] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45 (1):5－32.

[39] 刘敏,郎荣玲,曹永斌. 随机森林中树的数量[J]. 计算机工程与应用,2015,51(5):126－131.[LIU Min, LANG Rongling, CAO Yongbin. Number of trees in random forest[J]. Computer Engineering and Applications,2015,51(5):126－131.]

[40] DÍAZ-URIARTE R,DE ANDRES S A. Gene selection and classification of microarray data using random forest[J]. BMC Bioinformatics,2006,7(1):3.

[41] SVETNIK V,LIAW A,TONG C,et al. Random forest:A classification and regression tool for compound classification and QSAR modeling[J]. Journal of Chemical Information and Computer Sciences,2003,43(6):1947－1958.

Cropping structure extraction with *NDVI* time-series images in the northern Tianshan Economic Belt

XIONG Yuan-kang[1,2], ZHANG Qing-ling[1,3]

(1 *Research Center on Ecology and Environment of Central Asia*, *Xinjiang Institute of Ecology and Geography*, *Chinese Academy of Sciences*, *Urumqi* 830011, *Xinjiang*, *China*; 2 *University of Chinese Academy of Sciences*, *Beijing* 100049, *China*; 3 *School of Aeronautics and Astronautics*, *Sun Yat-sen University*, *Guangzhou* 510006, *Guangdong*, *China*)

Abstract: Limited water resources is the major factor affecting sustainable development in arid areas, and the water resources in arid areas are mostly used for agricultural irrigation. Rapidly and accurately mapping cropping structure in arid areas can provide an important basis for optimizing water use in agriculture. In this study, the Northern Tianshan Economic Belt, Xinjiang, China was chosen as the study area, and a method to map the cropping structure in this region with multi-source remote sensing data based on the Google Earth Engine (GEE) cloud platform was proposed. The Sentinel-2 and Landsat 7－8 data were chosen as the remote sensing data sources to extract the cropping structure in the study area through the following steps. First, in order to simplify the cropping structure extraction process and minimize the impacts from non-crop vegetation, a cropland mask was constructed by using the maximum *NDVI* value and slope information throughout the year in the study area. Second, according to the phenology calendars of the main crops in the study area, the time-series data of the maximum *NDVI* value and the corresponding date was calculated with remote sensing data. Then, the 10 feature bands were constructed. Third, the 10 feature bands were masked with the cropland mask. Based on these data together with the field samples, the random forest classifier was applied to cropping structure extraction. The accuracy evaluation results showed the overall accuracy of the classification results in 2018 was 92.19%, and the Kappa coefficient was 0.883. In order to further verify the accuracy of the classification algorithm, the crop structure in the study area in 2017 was also extracted, the classification results showed that the planted area of cotton, corn and wheat in the study area were 5 270 km^2, 2 000 km^2 and 2 340 km^2 respectively in 2017, and then compared it with the results of statistical yearbook data in 2017. The relative accuracy of cotton, corn and wheat planted area were 86.53%, 77.54% and 86.19%, respectively.

Key words: remote sensing in agriculture; Google Earth Engine; random forest classification; crop structure extraction; oasis agriculture

chinaXiv:201909.00154v1